

Available online at www.sciencedirect.com**SciVerse ScienceDirect**

Procedia - Social and Behavioral Sciences 70 (2013) 469 – 475

Procedia
Social and Behavioral Sciences

Akdeniz Language Studies Conference 2012

The impact of the assessors' attitude on the learners' oral assessment score pollution

Mahdiah Shafipoor^{a*}, Farnaz Latif^b^a Department of English Language Translation, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran^b Department of English Language, Shahr-e-Rey Branch, Islamic Azad University, Tehran, Iran

Abstract

Assessment plays a central and important role in teaching and learning. Every model of the teaching-learning process requires that teachers base their decisions - instructional, grading, and reporting - on some knowledge of student attainment of and progress toward desired learning outcomes. In this regard, assessment has a multitude of interpretations and uses and plenty of factors can affect the validity of these interpretations and uses. This paper presents the impact of assessors' attitude on score pollution in the context of oral assessment.

© 2012 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).
Selection and peer-review under responsibility of ALSC 2012

Keywords: Assessment; attitude; score pollution; validity; pedagogical knowledge

1. Introduction

In recent educational era, language instructors are expected to perform rather cooperatively in different stages of pedagogical decision making such as assessment which particularly involves instructors' constant judging and evaluation of learners' progress and overall performances. Thus, assessment plays a central role in pedagogical evaluations (e.g. Wilson, 2000; Brown, 2004; Fulcher & Davidson, 2007). To be more specific, during 1990s, language testers were by large interested toward test developments in two related fields: educational measurement and language teaching (McNamara, 1997).

* Mahdiah Shafipoor. Tel.: +98-912-1365283

E-mail address: m.shafipoor@yahoo.com

According to Bachman (2000), in educational measurement, there has been a movement toward what has been referred to variously as ‘alternative’ (e.g. Herman *et al.*, 1992), ‘authentic’ (e.g. TESOL 2000; Alderson & Banerjee, 2001), or ‘performance’ assessment (e.g. Solano-Flores & Shavelson, 1997), which has been moved quickly by widespread dissatisfaction with standardized multiple-choice tests, on the one hand, and by a movement toward the development of standards-based assessment, on the other. In this regard, much of the work on performance assessments of language ability has been situated in the context of language teaching (e.g. Brown & Hudson, 1998). In addition, there has been an increased focus on the role of raters in the assessment process. Moreover, several studies have found differences in raters’ behavior as a function of factors such as raters’ background, native language and amount of prior training or experience, both for assessments of speaking (e.g. McNamara, 1996) and of writing (e.g. Weigle, 1998). Researchers have also begun to investigate how raters arrive at their ratings (e.g. Pollitt & Murray, 1996).

Although, most of the studies carried out so far in the world of assessment have largely focused on EFL/ESL contexts other than Business English, the present study targeted at assessing General Business English Learners’ overall oral performance. To this end, the researchers hypothesized that raters’ attitude might have a significant impact on the oral assessment of the learners and any deficiency in this regard may lead to low validity of the scores and consequently score pollution. The term “score pollution” which was introduced by Mesick (1987) was defined as *any influence that “contaminates” the accuracy of achievement test scores*. Although the term was initially introduced and applied for achievement test scores, the term has been adopted here for oral assessments of General Business English Learners’ test scores as well. Haladyna (1992) identified three sources of contamination and reviewed the research bearing the seriousness of each as (1) test preparation, (2) situational factors, and (3) external conditions. In this study, the external conditions with the focus centred on the raters’ attitude are taken into account.

1.1. Objectives

This study was run in the training department of SAPCO (a supplying automotive parts company) in Tehran. The main mission of holding classes in this company was upgrading the staffs’ knowledge of general business skills such as socializing, meeting, telephoning, presentation, and negotiation to enable them to perform fluently in practical situations in contact with the foreign delegates and clients. However, in the present research work, among all these skills, the socializing skill was taken into account. Thus, to ensure the quality of the evaluation techniques applied to assess the learners’ overall progress in the socializing courses, the impact of raters’ attitude on the learners’ oral performance assessment was focused on. Therefore, the following research questions were drawn:

RQ1. What are the raters’ attitude toward the oral assessment process?

RQ2. Does the raters’ attitude have an effect on the learners’ oral assessment scores?

2. Method

2.1. Participants

In this study, three Business English instructors who were at the same time the raters (two of them majored in TEFL, namely the 1st and 2nd raters, and one of them majored in English literature, namely the 3rd rater) took part. Each instructor had 3 classes. Then, there were 9 classes with total of 138 learners who attended a twenty-session ‘Effective Socializing Skills’ course. The classes were held in the training department at SAPCO (Supplying automotive parts company) in Tehran. The minimum proficiency level of the participant learners was Intermediate.

2.2. Instruments

In the present study, initially, a semi-structured interview was devised and responses of 10 business English instructors were transcribed and coded to know their attitudes toward the issue of the assessment in general and the application of the performance-based oral assessment for learners of socializing skills in Business English courses in particular.

Secondly, based on the results of the interview coding, a five-scaled Lickert questionnaire was devised to estimate the three participant raters' attitudes toward the application of the performance-based oral assessment for learners of socializing skills in Business English courses. This questionnaire was piloted on 30 Business English raters to measure its reliability coefficient

2.3. Procedure

The design of the present study was a mixed method which entailed a combination of an initial qualitative phase and a final quantitative phase. The first phase of the study was initially performed quantitatively through the employment of a semi-structured interview, proceeded by the data coding and finally, completed by conducting a five-scaled Lickert questionnaire.

The second phase, on the other hand was conducted to quantify the collected data in the first phase. This process was run through a descriptive research methodology. Consequently, the correlation coefficient among the three participant raters' attitudes based on their responses to the questionnaire (first phase) was estimated through the application of the Pearson Product Moment Correlation.

2.4. Assessment setting

The present study was conducted at the presence of three raters, one of whom was the learners' own instructor. The learners were also paired into three levels, namely Excellent Performer (A), Moderate Performer (B), and Poor Performer (C), according to their class participation performance and socializing skill proficiency level which was identified in advance by their own instructors.

The time allocated to each pair's performance was a quarter and the questions were formed in the style of the performance-based speaking tasks which were shown on a video screen in PowerPoint format. The tasks entailed a discussion question followed by a socializing role play. The learners were given 5 minutes to rehearse the tasks before their actual performance. The raters' evaluation checklist was prepared with reference to a list of specification presented by Ellis & Johnson (1994, p. 48).

3.1. Qualitative Phase

3.1.1. Interview

As a result of the first phase's coding of the semi-structured interview responses, the following check list was developed. As it is presented in table 1, all of the 10 interviewed instructors had common concerns considering the significant role of the *raters*, *task types* and *assessment setting* on the quality of the oral assessment scores.

Table 1. Coding the interviews with 10 business English instructors

Coding the interviews with 10 business English instructors	
•	Rater's educational background
•	Rater's familiarity with the learners
•	Rater's role as the main instructor
•	Rater's role as a course planner
•	Types of tasks designed for the oral assessment
•	Authenticity of the tasks
•	Learners' preparation time through pair work
•	Final score as the average score assigned by all the raters

3.1.2. Questionnaire survey

As a result of the second stage in the first phase of the data analysis, two discrepancies were found comparing the responses of the three raters. In other words, in most instances, the 1st and 2nd raters' responses demonstrated more common grounds; however, comparatively, the 3rd rater's responses indicated a clear divergence. The core convergences in the 1st and 2nd raters' responses are presented in the following list:

- Rater's background educational and pedagogical knowledge
- Rater's pedagogical experience in assessment, syllabus design, materials development, and etc.
- Rater's personality type
- Rater's attitude toward the learners

Table 2. Raters' attitude results

	Professional background	1 st rater	2 nd rater	3 rd rater
1	I think raters who have passed TTC "teacher training courses" for Business English courses are more eligible for assessing learners' oral performance in business courses.	4	4	3
2	I think more experienced instructors are less biased raters for oral assessment of business courses.	5	5	1
3	I think raters' experience in rating can affect the overall scoring of oral assessment.	4	4	2
Pedagogical aspects				
4	I think the raters who are the main instructors are eligible for assessing learners' oral performance in business courses.	3	3	2
5	I think developing performance-based assessment tasks can qualify the validity of the assessment.	3	4	2
6	I think the rater's pedagogical knowledge level can affect the overall assessment scoring.	5	5	3
Ethical aspects				
7	I think the rater's personality type affects his/ her decision making for assessing learners' oral performance in business courses.	4	5	3
8	I think the rater's (if to be the main instructor in the class as well) positive or negative attitude toward the learners can affect the oral assessment overall scoring.	4	4	3
9	I think the rater's attitude toward the system can affect the overall assessment scoring.	5	4	2
10	I think the formal behavior of the assessors at the time of oral assessment can affect the oral assessment.	1	1	3
Guide key: no idea = 1, strongly disagree= 2, disagree= 3, agree= 4, strongly agree= 5				

3.2. Quantitative Phase

3.2.1. Correlational Analysis

In order to measure the degree of relationship among the three raters' scores, a correlational study was run by using the *Pearson Product Moment Correlation*. The result is given in table 3.

Table 3. Correlation coefficients among the three raters

Variable	1 st Rater	2 nd Rater	3 rd Rater
1 st Rater	1	*	*
2 nd Rater	.88	1	*
3 rd Rater	.53	.58	1

As table 3 shows, the correlation between the first and second raters' scores is high. On the contrary, the correlations between the second and the third raters, as well as the first and the third raters are low.

3. Conclusion

Teaching involves assessment. In making decisions about lesson content and sequencing, about materials, learning tasks and so forth, teachers have to determine the strengths and weaknesses of the alternatives available to them. They make selections based on their experience, on their understandings of learning, language development and of language proficiency itself, together with what they consider to be most appropriate and in the best interests of those they teach (Rea-Dickins, 2004).

In recent decades, there has been a movement toward what has been referred to variously as 'authentic', 'alternative', or 'performance' assessment. In performance assessment, there has been an increased focus on the role of raters in the assessment process and his pedagogical knowledge and experience which can play a crucial role in the validity of the assessment. As it was discussed in this paper, the present study was carried out to investigate the role of the raters' attitude in the oral assessment of the General Business English courses.

In conclusion, as it was presented in the data analysis section, the 3rd rater who graduated in a non-TEFL major and came up with different pedagogical knowledge and experience compared with other two raters (the 1st and 2nd raters) who majored in TEFL and had almost the same pedagogical knowledge and experience, held completely different attitude in oral assessment. And as a result, her scores correlated almost low with the other two raters.

Consequently, the results suggest that raters should have similar rather relevant pedagogical knowledge and experience in order to decide efficiently and fairly on the learners' assessment in general and oral assessment in particular and lack of this knowledge can have a negative effect on assessment validity which will result in score pollution. This result goes in line with the work of Brown, 1995 and McNamara, 1996 which proved that raters' background pedagogical knowledge and experience can affect assessment validity.

References

- Alderson, J. C., & Banerjee, J. (2001). *Language teaching*. Cambridge: Cambridge University Press.
- Anderson, J. O. (1990). Assessing classroom achievement. *The Alberta Journal of Educational Research* 36, 1_3.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing* 17 (1) 1–42.
- Brown, H. D. (2004). *Language assessment: principles and classroom practices*. Longman.
- Brown, J. D & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly* 32(4), 653–75.
- Ellis, M., & Johnson, C. (1994). *Teaching Business English*. Oxford: Oxford University Press.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: an advanced resource book*. Routledge Taylor and New York.
- Haladyna, T. (1992). *Test Score Pollution: Implications for Limited English Proficient Students*. Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues: Focus on Evaluation and Measurement. OBEMLA, 1992.
- Herman, J. L., Aschbacher, P. R. & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Longman
- McNamara, T. F. (1997). Performance testing. In Clapham, C. and Corson, D., editors, *Encyclopedia of language and education*. Volume 7. Language testing and assessment. Dordrecht: Kluwer Academic, 131–39.
- Pollitt, A., & Murray, N. L. (1996). What raters *really* pay attention to. In Milanovic, M. and Saville, N., editors, *Performance testing, cognition and assessment*. Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press, 74–91.
- Rea-Dickins, P. (2004). Test Score Pollution: Implications for Limited English Proficient Students. *Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues: Focus on Evaluation and Measurement*. OBEMLA, 1992.
- Solano-Flores, G. and Shavelson, R. J. (1997). Development of performance assessment in science: conceptual, practical, and logistical concerns. *Educational Measurement: Issues and Practice* 16(3), 16–25.
- TESOL (2000). *Integrating the ESL standards into classroom practice: grades pre-K_12*. Alexandria, VA: TESOL Inc.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing* 15(2), 263–67.

Wilson, R. J. (2000). *A model of assessment-in-practice*. Paper presented at the Annual Conference of the Canadian Society for the Study of Education, Edmonton, Alberta.